

On the automatic identification of Unobserved Components Models [★]

Diego J Pedregal¹[0000–0003–4958–0969] and Juan R
Trapero²[0000–0002–5879–3133]

¹ Industrial Engineering Politecnico, University of Castilla-La Mancha, 13071 Ciudad Real, Spain

² Faculty of Chemical Sciences and Technologies, University of Castilla-La Mancha, 13071 Ciudad Real

Abstract. Automatic identification of time series models is a necessity once the big data era has come and is staying among us. This has become obvious for many companies and public entities that has passed from a crafted analysis of each individual problem to handle a tsunami of information that has to be processed efficiently, online and in record time. Automatic identification tools has never been tried out on Unobserved Components models (UC). This chapter shows how information criteria, such as Akaike's or Schwarz's, are rather useful for model selection within the UC family. The difficulty lies, however, on choosing an appropriate and as general as possible set of models to search in. A set too narrow would render poor forecast accuracy, while a set too wide would be highly time consuming. The forecasting results suggest that UC models are powerful potential forecasting competitors to other well-known methods. Though there are several pieces of software available for UC modeling, this is the first implementation of an automatic algorithm for this class of models, to the best of the authors knowledge.

Keywords: Unobserved Components models · State Space systems · Kalman filter · smoother algorithm · Maximum Likelihood.

1 Introduction

The era of big data is provoking a revolution in many research areas. Indeed, it can be said that, in the area of time series forecasting the effect is particularly dramatic. Nowadays, big masses of time series ought to be forecast in short periods of time. Take as an example Walmart with 5,000 stores throughout the US which forecasting needs amounts to 10 millions per second! ([24], p. 828). Therefore, at least in such contexts, the traditional crafted approach to identification one time series at a time must be replaced by automatic identification alternatives.

[★] This chapter will be published in Valenzuela O., Rojas F., Pomares H., Rojas I. (eds.), Contribution to Statistics, Springer (in press).

Automatic selection of models has received a great deal of attention in the time series literature. This interest extends from classical modeling techniques such as regression analysis, Exponential Smoothing, ARIMA, transfer functions, etc. ([27, 6, 13, 15]), to modern Big Data techniques such as Artificial Neural Networks, Support Vector Machines, etc. ([11, 12, 29]).

Though it is almost impossible to make an exhaustive list of all the proposed forecasting methods in the literature, a good guidance to this variety may be found in the results of predictive competitions [21]. The most common forecasting methods to today are Exponential Smoothing (ETS) and ARIMA methods.

- Exponential Smoothing methods remain the most widely used modeling technique in day to day business and industry since the fifties [5]. Given the success and the fact that it was proposed initially as a heuristic method, a major revision has taken place in the last 20 years that has dramatically changed the vision of these techniques [16, 2].
- The second method most used is ARIMA. ARIMA models have expanded since the seventies after the publication of the influential book by Box and Jenkins [4]. Various methods have been proposed for automatic identification [6, 16], being TRAMO (together with SEATS) probably the ARIMA automatic identification procedure most used worldwide in official statistical agencies.

In all this scientific landscape there is a family of models with applications in many branches of science with rather good results, which has been conspicuously ignored, namely the Unobserved Components models (UC, [10, 28, 23, 3]). UC models aim explicitly at decomposing a vector of time series on components with economic meaning, normally trend, seasonal and irregular, although it may also include other components, typically cycles and exogenous variables modeled as linear regressions, transfer functions or nonlinear relationships.

The UC models have not been tested yet in automatic modelling settings for many reasons. First, UC models were developed by engineers and brought into economics by academics, with little interest on disseminating them among practitioners [22]. Second, UC methods are rarely taught at the undergraduate level, limiting access to the wide public. Third, there is a widespread intuition that UC models have nothing to add to other methods (especially Exponential Smoothing, [7]). Fourth, UC models are generally identified by hand, without any attempt to develop any automatic identification procedure. Finally, software packages are scarcer than packages for other more standard techniques. Some complete alternatives are, for example, STAMP [18], SSfpack [19], and SSspace [26].

The methods described in this chapter fill this gap by introducing a procedure to automatically select optimal UC models among a wide range of possibilities. The methods are useful in forecasting terms, but other byproducts are the estimated optimal components (trend, seasonal, irregular) that may be useful for other common and useful operations in time series analysis, such as smoothing, signal extraction, seasonal adjustment, detrending, etc.

The chapter is organised as follows. Section 2 presents briefly the UC models in general and the range of possibilities for each component. Section 3 shows how the UCs are inserted in the general State Space framework. The automatic identification procedure is presented in section 4. Section 5 shows the method working in practice on three real life case studies and compared to other alternatives. Finally, section 6 concludes.

2 Unobserved Components Models

The UC models aims at decomposing a time series into meaningful components. The most common decomposition is shown in equation (1), where T_t , S_t and I_t stand for the trend, seasonal and irregular components, respectively.

$$z_t = T_t + S_t + I_t \quad (1)$$

There have been many approaches to deal with this decomposition, from which the *structural* approach set up in a State Space (SS) framework is the most widespread.

Structural methods specify directly the particular dynamical models for each component involved, for which an ample range of possibilities exists. In general, all components are assumed stochastic, trends should be non-stationary by definition, seasonal components should show some sinusoidal behavior, and irregular components are usually considered either white or colored noise. The particular models chosen in this chapter for each component steam from a long tradition (see, among others, [9, 10, 28, 3]).

2.1 Trend components

All trends considered are particular cases of the Generalised Random Walk (or Damped Trend, DT) model shown in equation (2), where T_t^* is referred to as the trend ‘slope’, $0 < \alpha \leq 1$, $\eta_{T,t}$ and $\eta_{T,t}^*$ are independent white noise sequences with variances $\sigma_{\eta_T}^2$ and $\sigma_{\eta_T^*}^2$, respectively.

$$\begin{bmatrix} T_{t+1} \\ T_{t+1}^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} T_t \\ T_t^* \end{bmatrix} + \begin{bmatrix} \eta_{T,t} \\ \eta_{T,t}^* \end{bmatrix} \quad (2)$$

This model subsumes the following particular cases: i) Random Walk (RW), by eliminating the second equation (i.e., $T_{t+1} = T_t + \eta_{T,t}$ or setting $\alpha = 0$, $\sigma_{\eta_T^*}^2 = 0$ and $T_1^* = 0$); ii) Integrated Random Walk (IRW) with $\alpha = 1$ and $\sigma_{\eta_T}^2 = 0$, (it is equivalent to the well-known Hodrick-Prescott filter, [14, 28]); iii) Local Linear Trend (LLT) with $\alpha = 1$, see e.g., [10, 25, 3]. All these trends are stochastic and have at least one unit root ensuring they are not stationary.

2.2 Seasonal components

From all possibilities available in the literature (e.g., [9, 10, 28]), seasonal components used in this chapter take a stochastic trigonometric form. The seasonal component is built as the sum of individual sinusoidal terms for the fundamental period s and its harmonics. The number of harmonics in general is $\lceil s/2 \rceil = s/2$ for even s numbers, and $\lceil s/2 \rceil = (s - 1)/2$ for uneven s numbers.

The overall seasonal component is the sum of all the sinusoidal harmonics $S_{j,t}$ in equation (3), where $\omega_j = 2\pi j/s$ is the frequency of each harmonic, $S_{j,t}^*$ is an additional state necessary for the specification, and $\eta_{j,t}$ and $\eta_{j,t}^*$ are independent white noises with common variance σ_j^2 .

$$\begin{aligned} S_t &= \sum_{j=1}^{\lceil s/2 \rceil} S_{j,t} \\ \begin{bmatrix} S_{j,t+1} \\ S_{j,t+1}^* \end{bmatrix} &= \begin{bmatrix} \cos\omega_j & \sin\omega_j \\ -\sin\omega_j & \cos\omega_j \end{bmatrix} \begin{bmatrix} S_{j,t} \\ S_{j,t}^* \end{bmatrix} + \begin{bmatrix} \eta_{j,t} \\ \eta_{j,t}^* \end{bmatrix} \end{aligned} \quad (3)$$

An usual assumption regarding the seasonal component is to make all the variance noises equal to each other, i.e., $\sigma_j^2 = \sigma^2$, $j = 1, 2, \dots, \lceil s/2 \rceil$. This is indeed the case of the popular Basic Structural Model (BSM) of [10], but also of exponential smoothing models [15]. A much more flexible assumption is allowing all variances to be different (strictly as they are specified in equation (3)), an option that, though increasing the number of parameters, it renders models that are still feasible for most time series.

2.3 Irregular components

The empirical evidence in many cases is that, after taking into account trends and seasonal components, the remainder is just white noise. Therefore, the standard irregular component is just a Gaussian white noise with zero mean and constant variance σ_I^2 . However, for the cases where a serial correlation problem still remains, colored irregular components may be considered in the form of stationary ARMA(p, q) models, in general of low orders.

An ARMA(p, q) model is of the form

$$I_t = \phi_1 I_{t-1} + \phi_2 I_{t-2} + \dots + \phi_p I_{t-p} + \eta_{I,t} + \theta_1 \eta_{I,t-1} + \theta_2 \eta_{I,t-2} + \dots + \theta_q \eta_{I,t-q}$$

where $\eta_{I,t}$ is a Gaussian white noise with constant variance σ_I^2 , and ϕ_i ($i = 1, 2, \dots, p$) and θ_k ($k = 1, 2, \dots, q$) are unknown parameters that ought to be estimated from the data.

3 State Space systems

Once the model for all the components are specified, the *structural* UC approach proceeds by assembling all of them in a single linear Gaussian SS system by block concatenation of the individual models, in which equation (1) plays the role of the

observation equation. Then, all the statistical theory applicable to SS systems apply to the UC models straight away.

The minimum linear Gaussian SS system to deal with the whole set of models implemented in this chapter is shown in Equation (4).

$$\begin{aligned} \text{Transition equation: } \alpha_{t+1} &= \Phi\alpha_t + R\eta_t, \quad \eta_t \sim N(0, Q) \\ \text{Observation equation: } z_t &= Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, H) \end{aligned} \quad (4)$$

In these equations z_t is a univariate time series; α_t is a non-observable state vector of length n ; η_t and ϵ_t are the state and observational independent noises with zero mean Gaussian noises, with dimensions $r \times 1$ and 1×1 , respectively; the initial state vector is assumed to be stochastic with Gaussian distribution, i.e., $\alpha_1 \sim N(a_1, P_1)$, and independent of all data and noises involved in the system; the remaining elements in (4) are the so called system matrices with appropriate dimensions.

Much more complicated systems are possible, i.e., multivariate systems with time varying system matrices, non-linear, non-Gaussian, etc., but are not necessary in the present chapter and therefore are not considered here. For another toolbox with many of such capabilities see [26].

The main objective of SS systems is to obtain optimal estimations of the state vector and their covariances, in the sense of minimizing the mean square error, conditional on the particular model specified and all information available. Two sort of estimates are most common in practice:

- Filtered output by the well-known Kalman Filter. It provides the optimal state vector estimation using all the information available up to any point in time.
- Smoothed output by Fixed Interval Smoother algorithms, that renders the optimal estimates of the state vector based on the whole sample (past and future values), in a similar way to moving averages.

There are many issues related to state and parameter estimation in SS systems. The main ones concerning this chapter are:

- Missing data: they are naturally interpolated by the Kalman Filter and Smoother algorithms, because of their inherent recursive nature. Forecasts are also naturally produced by signaling the future values as missing data.
- The typical problem of initial conditions common to all dynamic systems is solved by using the exact initialization proposed by [3], known as diffuse filtering and smoothing.
- Model parameters scattered along the system matrices are estimated by maximising the diffuse log likelihood [3].
- Maximization of the log likelihood function requires optimization algorithms, which are usually Quasi-Newton type. Such algorithms take advantage of gradients of the likelihood surface, which may be computed either numerically or analytically [3]. Analytical gradients are possible for models that depend only on variance parameters in matrices Q and H in equation (4).

4 Automatic forecasting algorithm for UC

The automatic forecasting algorithm proposed below is based in information criteria (similar to [6, 16]) and performs remarkably well in practice, as will be shown in later worked examples. This is the first time that an algorithm of this nature is proposed in the literature about UCs.

The algorithm proceeds along the following steps:

- Step 1: Variance transformation. Decide whether to use the Box-Cox transformation or not [1]. This step is left to the user discretion because its benefits in terms in forecasting accuracy are not clear [15]. The approach by [8] is preferred here because it is not model dependent.
- Step 2: Model selection. A battery of models are estimated and the best is chosen according to the minimization of any information criterion, either the Akaike's (AIC) or Schwarz's (SBC), i.e.,

$$AIC = -2\ln(L^*) + 2k$$

$$SBC = -2\ln(L^*) + \ln(T)k$$

where L^* is the likelihood value at the optimum, T is the length of the time series and k the number of parameters in the model.

The set of models to search for are 23 and are all the possible combinations of trends (none, RW, LLT, DT), seasonal components (none, all harmonics with equal variance, all harmonics with different variances) and irregulars (none or Gaussian noise). The none trend / none seasonal / none irregular is excluded from the models set.

- Step 3: ARMA model selection. A low order ARMA model is then identified by AIC or SBC, on the innovations of the previous model estimated in step 2. The algorithm used is a simplified version of [16] for full non-stationary and seasonal ARIMA models. The simplification consists of searching exclusively on stationary and non-seasonal models, since both non-stationarity and seasonality are already captured by the trend and seasonal component.
- Step 4: Joint final estimation. If an ARMA model is detected in the previous step, then the full UC model with the ARMA irregular component embedded should be estimated.
- Step 5: Forecasting step. Final forecasts are produced with the best of models in steps 2 or 4, depending on which one exhibits the smallest information criterion value.

5 Case studies

The case studies considered below show how the UC automatic methods described in previous sections perform with respect to ARIMA and exponential smoothing (ETS) as implemented in the package `forecast` in R (functions `auto.arima` and `ets` were used, respectively) [16]. This package has gained the

role of a standard to which any new method may be confronted. Two further methods are added to make comparisons more comprehensive, namely a seasonal naïve method as a benchmark and the mean of the UC, ETS and ARIMA (see e.g., [20] about the importance of forecast combinations). Another dimension added to the case studies is checking whether the variance Box-Cox transformation improves forecast accuracy [1].

The case studies have been selected to be as varied as possible, they comprise a weather time series, another from macroeconomics and a the demand database of a retail business typical of supply chain applications. The sampling intervals of the time series are also varied, ranging from quarterly to daily.

Comparisons are carried out on the basis of two error metrics, the symmetric Mean Absolute Percentage Error (sMAPE) and the Mean Absolute Scaled Error (MASE), see equations (5), (6) and [17, 20]. There z_t and \hat{z}_t are the actual and forecast values at time t , respectively; T is the forecast origin; h is the forecast horizon; and n is the number of observations in the fitting sample

$$\text{sMAPE}_h = h^{-1} \sum_{i=1}^h \frac{2 |z_{T+i} - \hat{z}_{T+i}|}{|z_{T+i}| + |\hat{z}_{T+i}|} \times 100 \quad (5)$$

$$\text{MASE}_h = h^{-1} \sum_{i=1}^h \frac{|z_{T+i} - \hat{z}_{T+i}|}{(n-1)^{-1} \sum_{r=2}^n |z_r - z_{r-1}|} \quad (6)$$

5.1 Monthly average temperatures in Madrid at El Retiro weather station

Monthly average temperatures in Madrid from 1988 is shown in Figure 1. Data is compiled from the El Retiro weather station and are publicly available at www-2.munimadrid.es/CSE6/control/seleccionDatos?numSerie=14020000020.

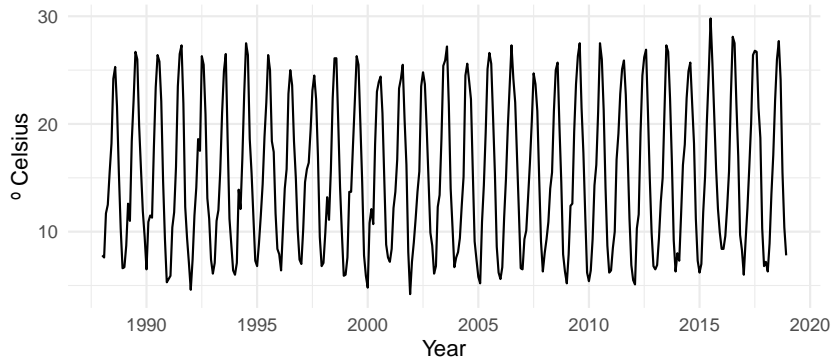


Fig. 1. Average temperatures in Madrid central

The series is dominated by the seasonal pattern. Maybe some noise is also present, but the trend, if any at all, is rather mild. All the methods are applied in a forecasting exercise consisting of a rolling out experiment in which the initial forecasting origin is set at December 2002 and the forecasting horizon is fixed at 12 months ahead. Then, one monthly observation is added at each iteration and the whole process is repeated until the end of the sample is reached. Therefore, 181 total rounds of 12 months-ahead forecasts from all models are produced and averaged along all the rounds to make final comparisons.

The models selected by the automatic identification of UCs confirm the initial intuitions based on Figure 1: trends are always either non-existent or damped with very small damping factors that effectively are very close to a non-existent trend (i.e., the α parameter in equation (2) always estimated smaller than 0.3); seasonal components are very strong, about half of the runs with common variance for all harmonics and half with different variances; the irregulars are identified either as non-existent (74% of all runs) or white noise (26%). The λ parameter of the Box-Cox variance transformation is in general close to 1, implying that the series does not exhibit heteroskedasticity problems.

Table 1. Error metrics for Madrid average temperatures error forecasts for several models and forecasting horizons. sMAPE is on the left part of the table and MASE on the right part. Minimum of each row is emphasized both for sMAPE and MASE.

h	sMAPE					MASE				
	Naïve	ETS	ARIMA	UC	Mean	Naïve	ETS	ARIMA	UC	Mean
1	12.649	9.171	8.688	8.560	8.736	1.083	0.799	0.740	0.742	0.753
2	12.706	9.247	9.009	8.876	8.976	1.086	0.803	0.770	0.768	0.774
3	12.732	9.321	9.104	9.004	9.084	1.088	0.809	0.779	0.779	0.784
4	12.755	9.357	9.153	9.073	9.140	1.091	0.812	0.784	0.785	0.789
5	12.784	9.384	9.192	9.111	9.177	1.094	0.814	0.788	0.789	0.792
6	12.804	9.393	9.213	9.146	9.200	1.097	0.815	0.790	0.792	0.794
7	12.814	9.382	9.224	9.161	9.208	1.098	0.814	0.791	0.793	0.795
8	12.821	9.376	9.233	9.171	9.216	1.100	0.813	0.792	0.794	0.795
9	12.825	9.387	9.233	9.176	9.224	1.100	0.814	0.792	0.794	0.796
10	12.833	9.387	9.239	9.180	9.229	1.101	0.813	0.792	0.794	0.796
11	12.837	9.387	9.249	9.187	9.237	1.101	0.813	0.792	0.795	0.796
12	12.844	9.384	9.250	9.197	9.240	1.102	0.813	0.792	0.795	0.796

The average forecasting performance of all models used are shown in Table 1, sMAPE at the left hand side and MASE at the right. Several conclusions may be extracted from this table: i) forecasts deteriorate with the horizon for all models, as expected; ii) all models show significant performance improvements over the Naïve, implying that they are really capturing the structure of the data beyond a naïve seasonal pattern; iii) the ordering of models from best to worst according to sMAPE is UC-Mean-ARIMA-ETS; iv) almost the same classification is produced with the MASE, except that ARIMA is the best and

UC is relegated to the second position, even though ARIMA, UC and Mean look actually very close to each other.

5.2 Spanish Gross Domestic Product (GDP)

Figure 2 shows the Spanish quarterly GDP between the first quarter of 1995 and the third quarter of 2019 in real terms (chain linked volume index). It follows a pattern similar to many western economies with a strong trend and seasonality and a big drop due to the 2008 recession followed by a final recovery.

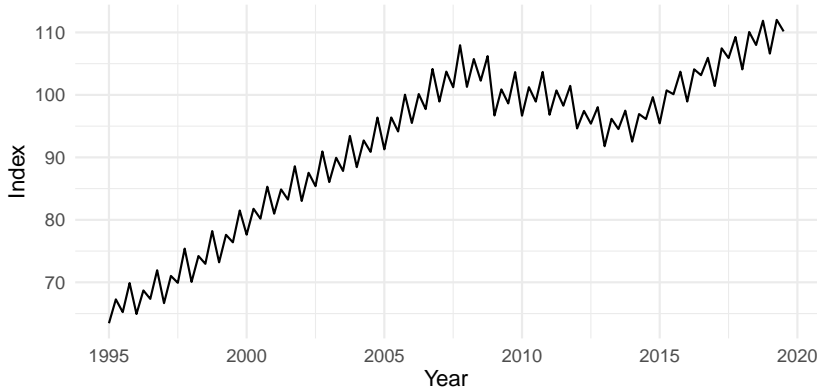


Fig. 2. Spanish real quarterly GDP between 1995 and 2019

The rolling out exercise in this case starts at the last quarter of 2009 and the forecasting horizon is fixed at 8 quarters ahead, i.e., the total number of 8 quarters-ahead forecast rounds are 32.

The UC model selected for most of the forecasting origins consists of a damped trend (with a damping parameter oscillating between 0.89 and 0.95), a seasonal component with equal variance for all harmonics and no irregular component. The estimated components for the whole sample may be seen in Figure 3.

The average forecasting performance of all models used are shown in Table 2. The table is divided in four quadrants reporting the average SMAPE and MASE metrics for each model with and without the variance Box-Cox transformation for horizons ranging from 1 to 12 months.

Table 2 offers some interesting insights into this forecasting exercise, some in common with the previous case study. Firstly, forecasts deteriorate with the horizon for all models. Secondly, all models show significant performance improvements over the Naïve. Thirdly, the absolute winner in this case regardless of the error metric is the UC method, followed by Mean, ETS and ARIMA.

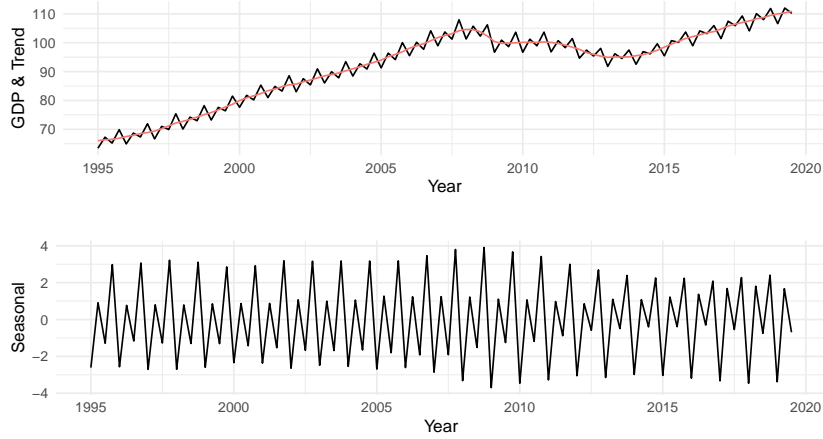


Fig. 3. Spanish real quarterly GDP between 1995 and 2019

Finally, this classification, with just a few exceptions, is independent of whether the Box-Cox transformation is used or not, the error metric and the forecasting horizon.

5.3 Demand database

The last case study is more complex than the previous ones, because it consists of all the daily demand time series, 142 in total, collected from a Spanish fresh food franchise. The series are available for the last 200 days and have a variety of properties in terms of predominance of components, volatility, etc. Two typical examples are shown in Figure 4. The bottom panel shows a time series dominated by the weekly pattern with a more or less stable mean, while the series at the top exhibits both a seasonal component and a decreasing trend much less stable.

The rolling experiment in this example consists of 48 runs for each of the 142 time series starting at day 140 and choosing a forecasting horizon of 14 days.

The heterogeneity of this bunch of time series is reflected in the variety of UC models automatically identified: trends are either damped (with damping factor varying between values close to 0 and 0.83) or Random Walks in equal parts (with a few of them non-existent); 11% of seasonal components are identified with different variances for each of the three harmonics, 23% of the series are estimated without any seasonal component and the rest are chosen as seasonal components with common variances; the irregular is non-existent in 22% of cases, while the rest are just white noise. Heterogeneity is also detected on the Box-Cox transformation that oscillates between -0.35 and 1 , with only 9 cases above 0.88 .

Table 3 summarises the values of the error metrics in a similar format to previous tables. In this case, only some selected forecasting horizons are shown

Table 2. Error measurements on Spanish GDP forecasts for several models and forecasting horizons. sMAPE is on top half and MASE at bottom. Results with Box-Cox variance transformation at the right side. Minimum of each row are emphasised.

sMAPE						sMAPE Box-Cox				
h	Naïve	ETS	ARIMA	UC	Mean	ETS	ARIMA	UC	Mean	
1	2.059	0.666	0.677	0.576	0.584	0.658	0.664	0.585	0.600	
2	2.099	0.781	0.785	0.666	0.688	0.754	0.788	0.665	0.699	
3	2.134	0.937	0.968	0.798	0.836	0.904	0.957	0.806	0.847	
4	2.164	1.112	1.134	0.943	0.995	1.066	1.128	0.954	1.000	
5	2.560	1.339	1.378	1.148	1.209	1.288	1.359	1.166	1.216	
6	2.848	1.562	1.613	1.353	1.414	1.499	1.593	1.381	1.417	
7	3.072	1.777	1.865	1.572	1.640	1.707	1.844	1.591	1.638	
8	3.255	1.990	2.086	1.781	1.842	1.915	2.075	1.803	1.844	

MASE						MASE Box-Cox				
h	Naïve	ETS	ARIMA	UC	Mean	ETS	ARIMA	UC	Mean	
1	0.773	0.245	0.246	0.210	0.214	0.241	0.241	0.213	0.219	
2	0.789	0.288	0.285	0.244	0.252	0.276	0.286	0.243	0.256	
3	0.803	0.344	0.352	0.293	0.307	0.331	0.349	0.296	0.310	
4	0.815	0.408	0.412	0.346	0.364	0.390	0.411	0.349	0.366	
5	0.965	0.491	0.499	0.422	0.442	0.471	0.494	0.427	0.445	
6	1.075	0.573	0.585	0.497	0.517	0.547	0.579	0.505	0.517	
7	1.160	0.651	0.675	0.578	0.598	0.622	0.670	0.582	0.597	
8	1.229	0.729	0.755	0.655	0.671	0.698	0.754	0.659	0.671	

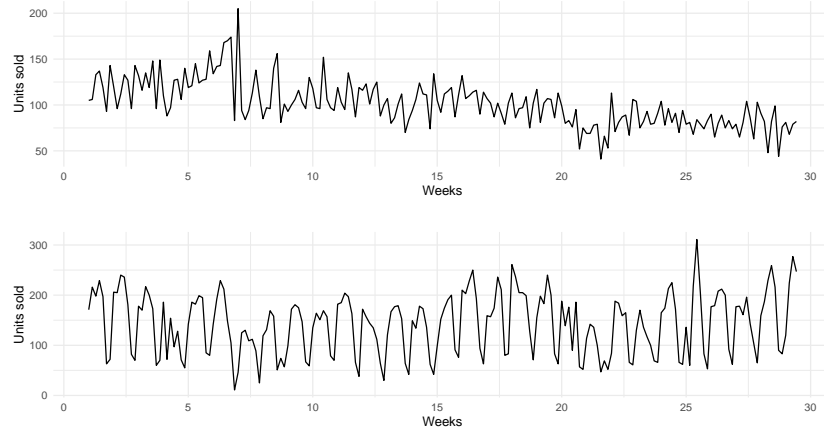


Fig. 4. Two examples of daily sales of a retailer in Spain.

to make the table shorter and the averages are calculated along time series and forecast origins, i.e., each value on the table is the average of $48 \times 142 = 6816$ forecast errors.

Table 3. Error measurements on demand time series for several models and selected forecasting horizons. sMAPE is on top half and MASE at bottom. Results with Box-Cox variance transformation at the right side. Minimum of each row are emphasised.

sMAPE						sMAPE Box-Cox				
h	Naïve	ETS	ARIMA	UC	Mean	ETS	ARIMA	UC	Mean	
1	6.016	4.717	4.914	4.710	4.634	4.684	4.865	4.696	4.619	
3	5.927	4.865	5.049	4.857	4.779	4.844	4.992	4.841	4.752	
5	5.911	4.947	5.084	4.931	4.841	4.942	5.041	4.926	4.825	
7	5.857	4.927	5.054	4.900	4.819	4.937	5.022	4.911	4.815	
9	5.833	4.927	5.051	4.887	4.813	4.940	5.022	4.909	4.812	
11	5.817	4.935	5.053	4.888	4.813	4.949	5.030	4.913	4.817	
13	5.814	4.940	5.059	4.891	4.818	4.959	5.038	4.920	4.826	
14	5.813	4.939	5.060	4.886	4.817	4.958	5.043	4.919	4.829	

MASE						MASE Box-Cox				
h	Naïve	ETS	ARIMA	UC	Mean	ETS	ARIMA	UC	Mean	
1	1.136	0.838	0.893	0.840	0.828	0.827	0.897	0.835	0.828	
3	1.115	0.871	0.924	0.873	0.860	0.858	0.922	0.866	0.856	
5	1.109	0.890	0.938	0.891	0.877	0.881	0.936	0.887	0.875	
7	1.097	0.894	0.939	0.894	0.880	0.890	0.939	0.894	0.881	
9	1.098	0.902	0.948	0.900	0.887	0.900	0.948	0.903	0.890	
11	1.094	0.906	0.951	0.902	0.890	0.910	0.952	0.909	0.896	
13	1.092	0.910	0.954	0.905	0.893	0.960	0.955	0.911	0.914	
14	1.092	0.912	0.956	0.906	0.894	0.962	0.958	0.913	0.916	

Table 3 shows that all methods outperform the Naïve method. Forecasts roughly worsen for longer forecasting horizons (not so clear as in previous case studies). The winner method is unambiguously the combination of methods (Mean), followed by UC, ETS and ARIMA. There is an interesting distinct behaviour of error depending on the horizon, because for horizons up to 7 days ahead (there are some variations depending on the method) the Box-Cox transformation gives lower errors that forecasts with no transformation. There is only one exception to the previous rule, the sMAPE metric for the ARIMA method, for which forecasts are always better with Box-Cox transformation.

6 Conclusions

This chapter presents a novel automatic identification procedure for UC models, consisting of estimating a wide range of possible models and selecting the best according to any information criterion, like Akaike's or Schwarz's. This sort of algorithm is pretty useful in Big Data contexts where many time series ought to be processed reliably in rather fast times.

The most important point is choosing an appropriate set of UC models, wide enough to be able to represent efficiently as many time series as possible. In that regard, the trend components available are either none, Random Walk, Local

Linear Trend or Damped Trend (see equation (2)). The seasonal component are either none, seasonal harmonics with equal variances or with different variances. Finally, irregulars are allowed to select among none, white noise or ARMA processes. As far as the authors are concern, this is the widest set of UC models available in the literature.

The previous algorithm is assessed on three case studies in comparison with other well-known methods, namely ARIMA and ETS as implemented in the `forecast` package in R.

The results show that the proposed identification algorithm is strongly competitive with the rest, being the best very often. Apart from this general conclusion that is the most important, there are other findings that were not specifically pursued, and therefore should be considered only partial to the particular case studies included. Firstly, there is little disagreement between both error metrics (sMAPE and MASE) when ordering the forecasting methods. Secondly, there are not clear improvements in forecasting accuracy when the Box-Cox variance transformation is used. Finally, combination of forecasts (at least the mean used in this chapter) does not imply better forecasts, only in the last case study the combination outperformed the rest.

To sum up, UC models automatically identified provides a nice tool that may enter the forecasters toolbox, with some nice by-products consisting of the optimal decomposition of time series in trend, seasonal component and irregular, that often are required for detrending, signal extraction, seasonal adjustment, etc.

Acknowledgements

This work was supported by the Spanish Government (MINECO/FEDER, UE) under the project with reference DPI2015-64133-R and by the Vicerrectorado de Investigación y Política Científica from UCLM through the research group fund program (PREDILAB; DOCM 24/01/2019 [2019/536]) and by the European Regional Development Fund.

References

1. Box, G., Cox, D.: An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**, 211–252 (1964)
2. De Livera, A., Hyndman, R., Snyder, R.: Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* **106**, 1513–1527 (2011)
3. Durbin, J., Koopman, S.: *Time series analysis by state space methods*. Oxford University Press (2012)
4. G, B., G, J., G, R., G, L.: *Time series analysis: forecasting and control*. John Wiley & Sons (2015)
5. Gardner, E.: Exponential smoothing: The state of the art – part ii. *International Journal of Forecasting* **22**, 637–666 (2006)

6. Gómez, V., Maravall, A.: Automatic Modeling Methods for Univariate Series. In: A Course in Time Series, pp. 171–201. John Wiley & Sons, Inc. (2001)
7. Gooijer, J., Hyndman, R.: 25 years of time series forecasting. *International Journal of Forecasting* **22**, 443–473 (2006)
8. Guerrero, V.M.: Time-series analysis supported by power transformations. *Journal of Forecasting* **12**(1), 37–48 (1993)
9. Harrison, P., Stevens, C.: Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)* **38**, 205–247 (1976)
10. Harvey, A.: Forecasting, structural time series models and the Kalman filter. Cambridge university press (1989)
11. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference, and prediction. New York: Springer Publishing (2009)
12. Haykin, S.: Neural networks and learning machines. New Jersey: Prentice Hall (2008)
13. Hocking, R.: The analysis and selection of variables in linear regression. *Biometrics* **32**, 1–49 (1976)
14. Hodrick, R., Prescott, E.: Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* **29**, 1–16 (1997)
15. Hyndman, R., Koehler, A., Ord, J., Snyder, R.: Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media (2008)
16. Hyndman, R., Khandakar, Y.: Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software* **27**, 1–22 (2008)
17. Hyndman, R., Koehler, A.: Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**(4), 679 – 688 (2006)
18. Koopman, S., Harvey, A., Doornik, J., Shephard, N.: Structural Time Series Analysis, Modelling, and Prediction using STAMP. London: Timberlake Consultants Press (1999)
19. Koopman, S., Shephard, N., Doornik, J.: Statistical Algorithms for Models in State Space Form: SsfPack 3.0. Timberlake Consultants Press (2008)
20. Makridakis, S., Hibon, M.: The m3-competition: results, conclusions and implications. *International Journal of Forecasting* **16**, 451–476 (2000)
21. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting* **34**, 802–808 (2018)
22. Pedregal, D.: State space modeling for practitioners. *Foresight* **54**, 21–25 (2019)
23. Pedregal, D., Young, P.: Statistical approaches to modeling and forecasting time series. In: A Companion to Economic Forecasting, pp. 69–104. Blackwell Publishing Ltd (2002)
24. Seaman, B.: Considerations of a retail forecasting practitioner. *International Journal of Forecasting* **34**, 822–829 (2018)
25. Taylor, C., Pedregal, D., Young, P., Tych, W.: Environmental time series analysis and forecasting with the **Captain** toolbox. *Environmental Modelling & Software* **22**(6), 797–814 (2007)
26. Villegas, M., Pedregal, D.: Sspace: A toolbox for state space modelling. *Journal of Statistical Software* **87-5**, 1–26 (2018)
27. Young, P.: Recursive Estimation and Time-Series Analysis. Springer-Verlag (2011)
28. Young, P., Pedregal, D., Tych, W.: Dynamic harmonic regression. *Journal of forecasting* **18**(6), 369–394 (1999)
29. Zhang, G.: Times series forecasting using a hybrid arima and neural network model. *Neurocomputing* **50**, 159–175 (2003)